Earth Surface Temperature Data Correction and Validation Processes Using MATLAB

Anjeza Arapi (LaGuardia Community College of CUNY)
Karina Shah (The Dalton School)
Dana Hojnowski (Educator, North Bergen HS)
Mentors: Dr. Yasser Hassebo and Dr. Reginald Eze
yhassebo@lagcc.cuny.edu

Abstract

Remote sensing is the acquisition of information from a distance. Remote sensing instruments installed on satellites launched by NASA, along with many other international space agencies, can collect data on carbon dioxide, aerosols, biomass, temperature, and precipitation. Analysis of global maps of Earth Surface Temperature (EST) are required for many applications, including climate change research, global warming studies, and weather prediction. This paper focuses on the correction and validation of EST data received from satellites, in an effort to investigate global warming. In this research, 992 satellite EST data files (a total of 10,285,056 data points from four years) were provided. Anomalies and errors in the raw data are represented as '-1000'. To investigate the EST data, outliers were determined and corrected, data files were re-shaped, summarized, and analyzed statistically through five steps of averaging and validation processes using MATLAB. Through the five-step data correction process, data sets were expressed in histograms, contour maps, and tables of all calculated statistical parameters (mean, standard deviation (StD), skewness (Sk), and kurtosis). For validation, results of each step were compared with the corresponding statistical value of NASA as a reference for the same period of time. Results show the EST calculated statistical parameters shifted closer to NASA's EST reference values as anomalies in the raw data are replaced with zeros, and subsequently with NaNs as placeholders. The closest calculated statistical parameters to the NASA references occurred when the outliers were replaced with NaNs. Further data correction steps will be introduced and the global warming hypothesis will be analyzed and discussed.

Introduction

Engineers have developed instruments to detect the subtle differences in thermal radiation that real world objects emit. These sensors have been installed on satellites, which can collect data from areas otherwise impossible to reach. A thermal remote sensor measures the total radiant energy exiting an object and sends these to ground stations where scientists manipulate them using thermal radiation laws. Scientists calculate T_{rad} using the Stefan-Boltzman law, $E_{rad} = \sigma^* T_{rad}^4$, then convert T_{rad} to T_{true} using Kirchoff's law, $T_{true} = \varepsilon^{-1/4} * T_{rad}$ [5].

The sun is Earth's source of light. It emits a continuous spectrum, though its peak emissions lie in the visible band of light. The Earth absorbs this energy and re-emits it primarily as infrared radiation, which is referred to as the Earth's radiant flux. An object's radiant flux is directly proportional to its kinetic temperature, so we can use remote sensing to measure the Earth's radiant flux, which will allow us to obtain the temperature of its surface [5].

The data acquired by NASA contains thousands of errors due to anomalies on the surface of the Earth, and we must find and correct these anomalies. Using MATLAB, a high-level language and interactive environment for numerical computation, visualization, and programming, we completed four stages of data correction processes to analyze the EST [6].

Earth's surface temperature is particularly important because it allows us to track our planet's changing climate. A warming trend in our data through our correction and validation processes will provide evidence to support global warming.

Methods

Earth Surface Temperature Data Correction and Validation

Scientific Background

Mathematical Background

Matlab

Methods

(1) **Import** the data sets into MATLAB and concatenate three-dimensionally to ensure that the geographic orientation remains constant.
(2) **Count** the number of outliers in each data set.
(3) **Analyze** raw data (calculating the mean, standard deviation, skewness and kurtosis and graph the Histograms and the contour Maps).
(4) Replace all of the outliers with **zeros**. Re-analyze.
(5) Replace the zeros in the data set with **NaNs**. Re-analyze.
(6) Linear Interpolation: Replace outliers in the raw data set with the average of the closest non-negative values to the **left and right** of data points. Re-analyze.
(7) Interpolation: Replace outliers from original data set with the average of the closest non-negative values from **above, below, left, and right** of the data points. Re-analyze.
(8) Weighted Area correction: Take into account the latitude-longitude quadrangles in the Earth spherical geometry.

Conclusion

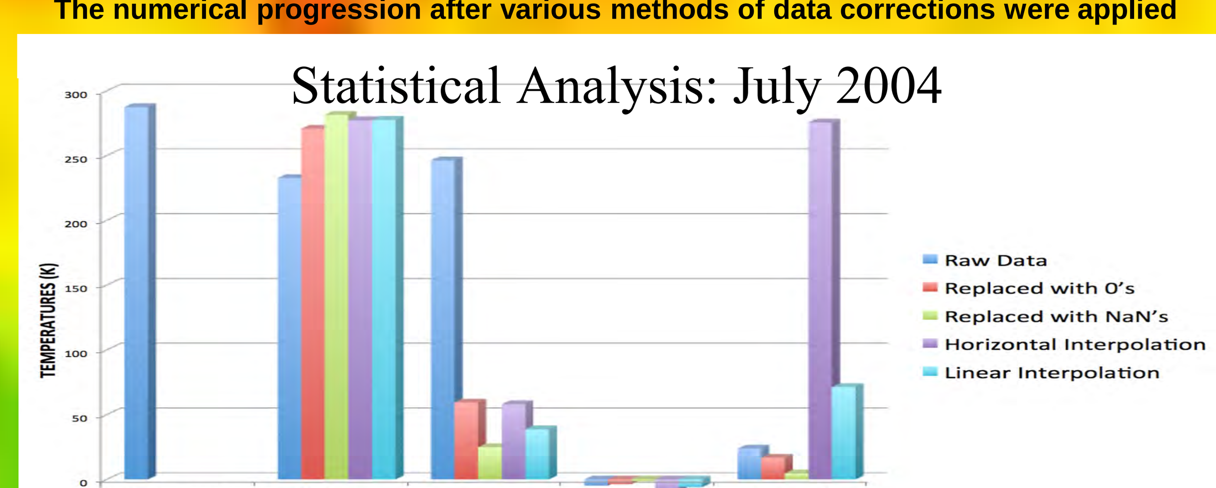
YES

NO

Results

	NASA Mean (°C)	Raw Data Mean (°C)	Zeros Mean (°C)	NaNs Mean (°C)	Horizontal Interp Mean (°C)	Linear Interp Mean (°C)	Weighted Area Mean (°C)
Jul-02	14.61	-94.28	-13.38	9.6	-13.25	-9.65	7.74
Jul-03	14.55	-37.91	-1.67	8.7	4.19	4.59	14.94
Jul-04	14.27	-40.5	-2.46	8.37	4.03	4.39	14.97
Jul-05	14.66	-49.85	-4.3	8.66	2.26	2.33	13.89

The numerical progression after various methods of data corrections were applied



The EST data was given under the assumption: all longitude lines (meridians) are vertical parallel lines and each latitude line is parallel to all others horizontally (Fig. a). The data files given in a matrix format. Each matrix/file represent the average of 3hr of EST (i.e., 8 files per day). The matrix size is 72 x 144. Each element in the matrix is the average EST for a specific part of the globe (Earth's grid is divided into equal areas, each part is 2.5 x 2.5 degrees).

Area Correction

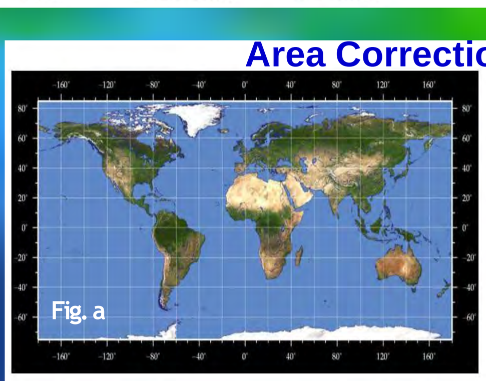


Fig. a

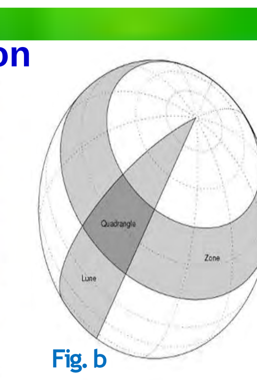
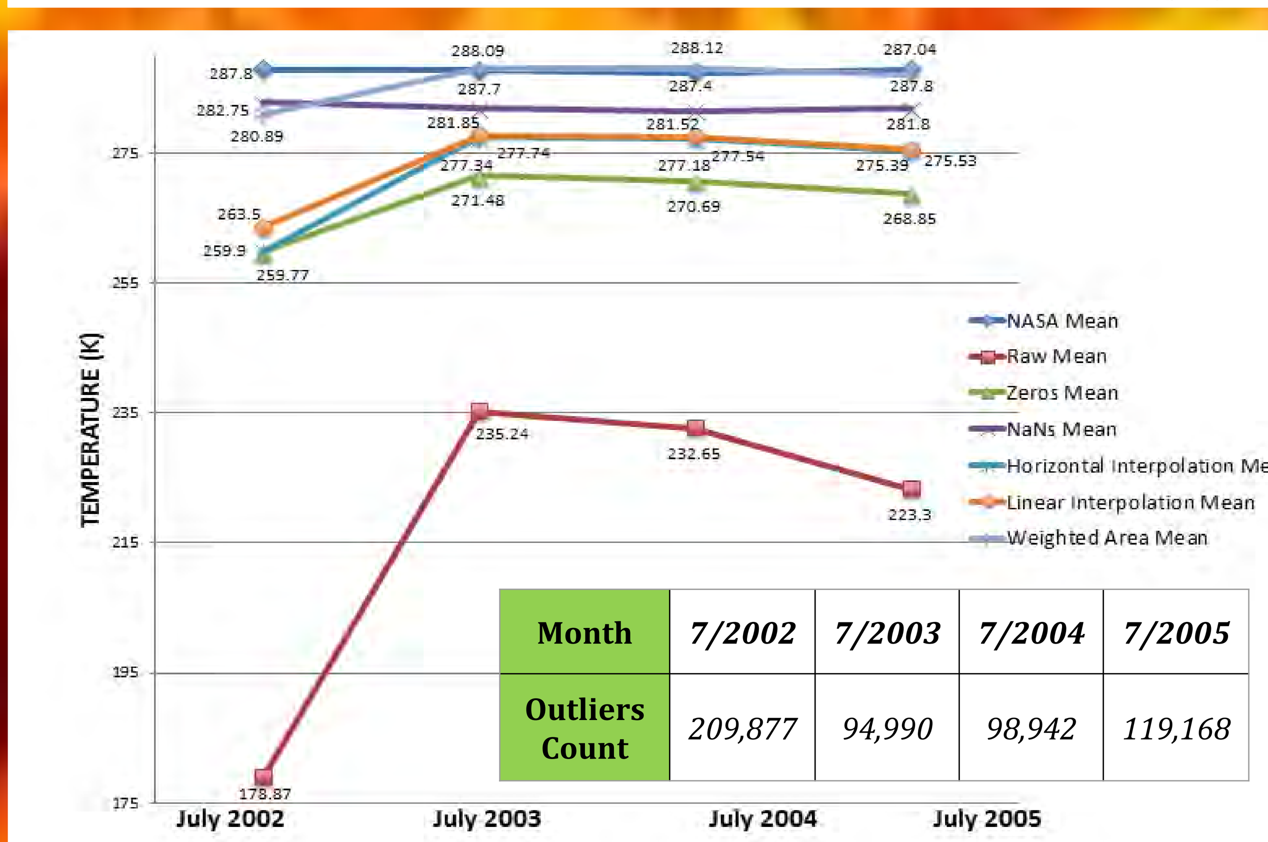


Fig. b

As shown in Fig. b, due to spherical shape of the Earth, longitude lines (meridians) are not parallel. The angle between meridians became smaller if one moved from the equator toward the pole(s). Taking this fact into account in our EST statistical calculation, that is called "Weighted Area", our calculated EST mean gets closer to the NASA's EST reference mean.

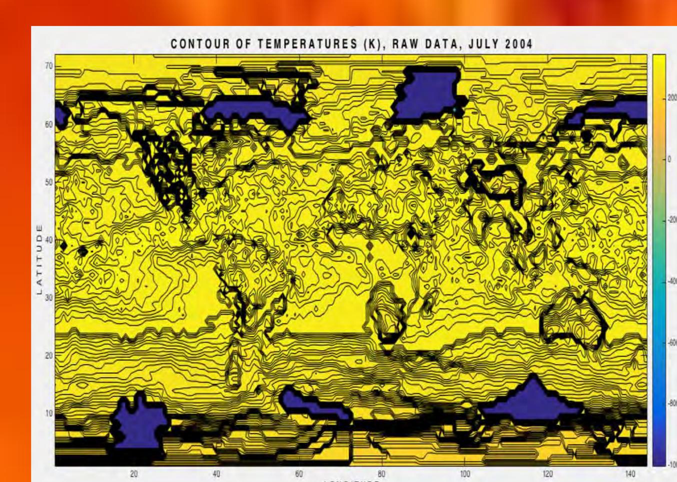
Results

2002-2005 Earth Surface Temperature (K)

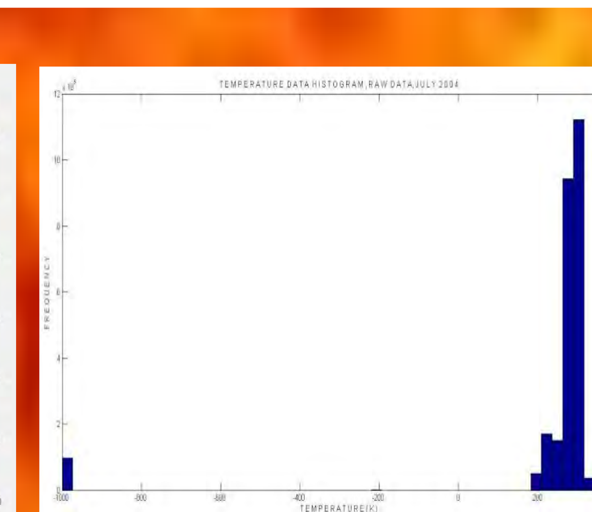


Month	7/2002	7/2003	7/2004	7/2005
Outliers Count	209,877	94,990	98,942	119,168

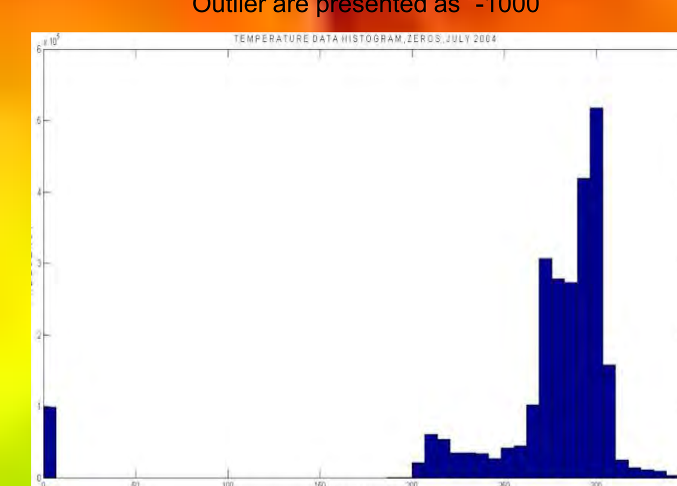
2004 Earth Surface Temperature Visuals



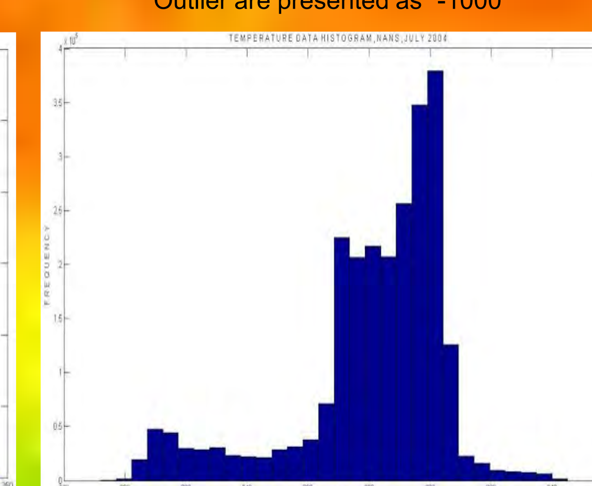
EST Contour, Raw Data, July 2004
Outlier are presented as "-1000"



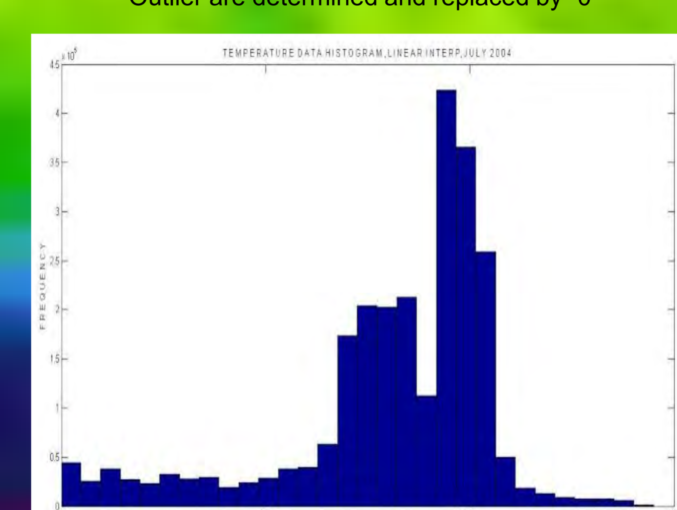
EST Histogram: Raw Data, July 2004
Outlier are presented as "-1000"



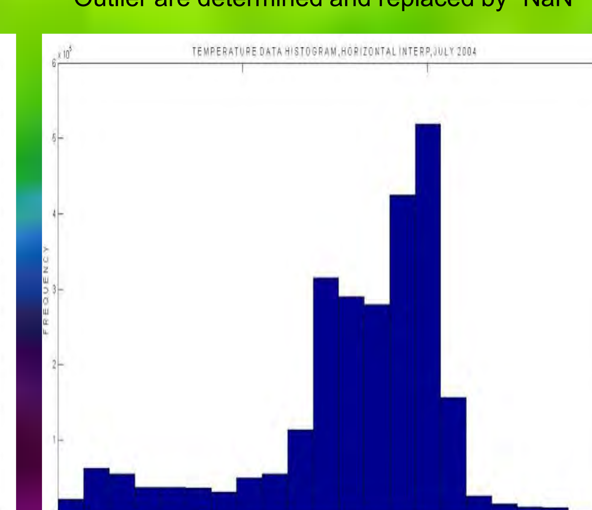
EST Histogram: July 2004
Outlier are determined and replaced by "0"



EST Histogram: July 2004
Outlier are determined and replaced by "NaN"



EST Histogram: July 2004
Outliers were estimated: Interpolation Stage



EST Histogram: July 2004
Interpolated data were re-calculated: Weighted Area Stage

Conclusion

Missing data due to obstacles such as cloud and/or pollutions, Instrumental and measurement errors, and unusual values (compared to the minimum and the maximum temperature in Earth and compared to the adjacent values) were replaced by "-1000" in the raw data and yielded outliers in the EST data set. Outliers caused bias in the statistical analysis results.

To improve the bias, several steps of data corrections were conducted. Each step of this data correction process improved the bias and shifted the mean towards the NASA's EST reference mean. Step 1) We had to locate these outliers and remove them. Removing outliers led to re-allocating EST values in wrong areas. Step 2) As a practice stage, outliers were replaced with zeros to fill the gaps and to keep the geometric grid of the earth map. These zeros improved the bias in the statistical analysis results by shifting the mean towards the NASA reference mean as shown in figure. Step 3) In this stage outliers were replaced with placeholders, known as NaN's. The analysis using nanmean, in Matlab, provided better statistical results compared to the previous stage because the nanmean is ignoring the number of Nan values when calculating the mean (that decreased the dominator) and shifted our calculated EST mean closer towards the NASA reference mean. Step 4) Estimating the outliers (Interpolations) using the adjacent non-negative values improved the bias , but not as good as step 3. Despite the results, step 4 represented the real situation to the EST. Step 5) The weighted area correction due to Earth spherical shape (Fig. a & b) of our interpolated data yielded EST mean closest to the NASA EST reference mean, because it reflects a more accurate model of the Earth.

On the other hand, comparing the average EST from 2002 to 2005, it becomes clear that there is no visible trend. Rather, the averages appear to be in flux, shifting down from 2002 to 2003, down again from 2003 to 2004 and up from 2004 to 2005. There is a lack of increase in global temperature over these four years; this span of time is inefficient to support global warming. The research results showed that our values deviated significantly from NASA's reference in July 2002, further research must be done to figure out this discrepancies involving more data and further investigations such as the impact of volcanic activity.

References

(1) Amos, J. (2013, December 9). Coldest Spot on Earth Identified by Satellite. Retrieved from http://www.bbc.com/news/science-environment-25287806.
(2) Cool Cosmos. (n.d.). Retrieved from http://coolcosmos.ipac.caltech.edu/cosmic_classroom/ir_tutorial/what_is_ir.html.
(3) Esri - GIS Mapping Software, Solutions, Services, Map Apps, and Data. (n.d.). Retrieved from http://www.esri.com/.
(4) "GLOBAL Land-Ocean Temperature Index in 0.01 Degrees Celsius Base Period: 1951-1980." *Data.giss.nasa.gov*. N.p., n.d. Web. http%3A%2F%2Fdata.giss.nasa.gov%2Fgistemp%2Ftabledata_v3%2FGLB.Ts%2BdSST.txt.
(5) Jensen, John R. (2007). *Remote Sensing of the Environment.* Upper Saddle River, NJ: Pearson Education, Inc.
(6) NASA. Retrieved from http://www.nasa.gov/.
(7) NOAA National Centers for Environmental Information, State of the Climate: Global Analysis for July 2004, published online August 2004, retrieved on July 29, 2015 from http://www.ncdc.noaa.gov/sotc/global/200407.
(8) The Sun and its Energy. (n.d.). Retrieved from http://photojournal.jpl.nasa.gov/catalog/PIA12098.
(9) University of Wisconsin-Madison Space Science and Engineering Center. (n.d.). Retrieved from https://cimss.ssec.wisc.edu/goes/blog/archives/600.
(10) West, Larry. "2005 Weather - 2005 - A Year of Extremes." N.p., n.d. Web. 29 July 2015.

Acknowledgements

National Aeronautics and Space Administration (NASA)
NASA Goddard Space Flight Center (GSFC)
NASA Goddard Institute for Space Studies (GISS)
NASA New York City Research Initiative (NYCRI)
Dept. of Education – The Alliance for Continuous Learning Environments in STEM (CILES) grant # P031C110158
LaGuardia Community College (LAGCC)

Dr. William Rossow, Distinguished Professor, EE Dept, CCNY
Dr. Jorge Gonzalez, ME Professor, PI of CILES grant, CCNY
Dr. John R.E. Toland, Physics Professor, LAGCC